

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/931,267

DATE: 08/23/2001

TIME: 17:30:55

Input Set : A:\N1422005.txt

Output Set: N:\CRF3\08162001\I931267.raw

ENTERED

3 <110> APPLICANT: Chapple, Clinton
 4 Franke, Rochus
 5 Ruegger, Maxwell
 7 <120> TITLE OF INVENTION: Genes Encoding P-Coumarate 3-Hydroxylase (C3H) and Methods
 of Use
 9 <130> FILE REFERENCE: N1422-005
 C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/931,267
 C--> 11 <141> CURRENT FILING DATE: 2001-08-16
 11 <150> PRIOR APPLICATION NUMBER: US 60/225554
 12 <151> PRIOR FILING DATE: 2000-08-16
 14 <160> NUMBER OF SEQ ID NOS: 7
 16 <170> SOFTWARE: PatentIn version 3.0
 18 <210> SEQ ID NO: 1
 19 <211> LENGTH: 1527
 20 <212> TYPE: DNA
 21 <213> ORGANISM: Arabidopsis thaliana
 23 <400> SEQUENCE: 1
 24 atgtcgtggt ttctaatagc ggtggcgaca atcgccgcgcg tcgtatccta caagctaate 60
 26 caacggctaa gatacaagtt cccaccaggc ccaagcccca agccgatcgt cggtaacctc 120
 28 tacgacataa aaccgggtccg gttcagatgt tactacgagt gggctcaatc ttatggacca 180
 30 atcatatcgg tctggatcgg ttcaattcta aacgtggtcg tatctagcgc cgagctagca 240
 32 aaagaagttc tgaaagaaca cgaccagaaa ctgcgcgacc ggcaccggaa cagatcgacg 300
 34 gaagcattta gccgcaacgg tcaggatcctt atatggggcc attatgggccc tcattacgtg 360
 36 aaggtgagaa aagtttgcac gcttgagctc ttcacaccga aacgactcga gtctctcaga 420
 38 cctatccgtg aagatgaagt caccgccatg gttgaatccg tcttcagaga ctgtaacctt 480
 40 cctgaaaaca gagcaaaagg tttaacaactg aggaagtact taggagcggg tgcgttcaac 540
 42 aacataacgc ggcagacccg ttggaagcgt tttatgaacg ctgaaggtgt tgtggacgag 600
 44 caagggtctg agttcaaggc catagtatcc aacgggtctga agctagggtgc ttcactgtca 660
 46 atagctgaac acatcccgtg gctcaggtgg atgtttccgg ctgatgagaa ggcgtttgct 720
 48 gagcacgggg ctgcgtcgtg ccgcctcact cgagctatca tggaggagca tactttggcc 780
 50 cgtcaaaagt ctagtggagc gaaacagcat ttcgttgatg cgttgctaac gttgaaggat 840
 52 cagtatgata ttagtgagga tactatcatt ggtcttctat gggatatgat cacggcaggg 900
 54 atggacacga cagcgataac agcggaatgg gcgatggcgg aaatgatcaa gaatccaaga 960
 56 gtgcaacaaa aagtgcaga agagttcgac agagtgggtg gacttgaccg gatcttaacc 1020
 58 gaggcagatt tctcccgtt accttacttg caatgcgtgg tgaaagagtc attcaggctg 1080
 60 catcctccaa cgcctctaag gctacctcac cgaagcaacg cagatgtcaa gatcggaggc 1140
 62 tatgatattc ccaaaggatc aaacgttcat gtgaatgtgt gggctgtggc tagagaccgc 1200
 64 gctgtatgga aaaatccatt tgagttttag ccagagagat tcttggaaga agatgttgac 1260
 66 atgaagggtc atgatttttag gctgcttccg tttggagctg gaagacgggt ttgtcccggg 1320
 68 gcacaacttg gtatcaattt ggtaacttcg atgatgagtc atttgcttca ccattttgtt 1380
 70 tggacacctc ctcaaggagc taaaccggag gagattgaca tgtctgaaaa ccctggactc 1440
 72 gttacttaca tgcgtacccc tgtgcaagcg gttgcaacgc ctcggttgcc ttcggatctg 1500
 74 tacaacgcg tgccttacga tatgtaa 1527
 77 <210> SEQ ID NO: 2
 78 <211> LENGTH: 1527
 79 <212> TYPE: DNA
 80 <213> ORGANISM: Arabidopsis thaliana
 82 <400> SEQUENCE: 2

RAW SEQUENCE LISTING

DATE: 08/23/2001

PATENT APPLICATION: US/09/931,267

TIME: 17:30:55

Input Set : A:\N1422005.txt

Output Set: N:\CRF3\08162001\I931267.raw

```

83 atgtcgtggt ttctaatagc ggtggcgaca atcgccgcgc tcgtatccta caagctaate 60
85 caacgggctaa gatacaagtt cccaccaggc ccaagcccca agccgatcgt cggtaacctc 120
87 tacgacataa aaccgggtccg gttcagatgt tactacgagt gggctcaatc ttatggacca 180
89 atcatatcgg tctggatcgg ttcaattcta aacgtggtcg tatctagcgc cgagctagca 240
91 aaagaagttc tgaaagaaca cgaccagaaa ctcgccgacc ggcaccggaa cagatcgacg 300
93 gaagcattta gccgcaacgg tcaggatctt atatgggccc attatgggccc tcattacgtg 360
95 aaggtgagaa aagtttgcac gcttgagctc ttcacaccga aacgactcga gtctctcaga 420
97 cctatccgtg aagatgaagt caccgccatg gttgaatccg tcttcagaga ctgtaacctt 480
99 cctgaaaaca gagcaaaagg tttacaactg aggaagtact taggagcggg tgcgttcaac 540
101 aacataacgc ggctagcctt tgggaagcgt tttatgaacg ctgaagggtg tgtggacgag 600
103 caagggcttg agttcaaggc catagtatcc aacgggtctga agctaggtgc ttcactgtca 660
105 atagctgaac acatcccgtg gctcagggtg atgtttccgg ctgatgagaa ggcgtttgct 720
107 gagcacgggg ctcgctcgtg ccgcctcact cgagctatca tggaggagca tactttggcc 780
109 cgtcaaaaagt ctagtggagc gaaacagcat ttcgttgatg cgttgctaac gttgaaggat 840
111 cagtatgata ttagtgagga tactatcatt ggtcttctat gggatatgat caccgcaggg 900
113 atggacacga cagcgataac agcggaatgg gcgatggcgg aaatgatcaa gaatccaaga 960
115 gtgcaacaaa aagtgcaga agagttcgac agagtgggtg gacttgaccg gatcttaacc 1020
117 gaggcagatt tctcccgtt accttacttg caatgcgtgg tgaaagagtc attcaggctg 1080
119 catctccaaa cgcctctaatt gctacctcag cgaagcaacg cagatgtcaa gatcggaggc 1140
121 tatgatattc ccaaggatc aaacggttcat gtgaatgtgt gggctgtggc tagagaccgc 1200
123 gctgtatgga aaaatccatt tgagtttaga ccagagagat tcttggaaga agatgttgac 1260
125 atgaagggtc atgattttag gctgcttccg tttggagctg gaagacgggt ttgtcccgtt 1320
127 gcacaacttg atatcaattt ggttaacttg atgatgagtc atttgcttca ccattttggt 1380
129 tggacacctc ctcaagggac taaaccggag gagattgaca tgtctgaaaa ccctggactc 1440
131 gttacttaca tgcgtacccc tgtgcaagcg gttgcaacgc ctcggttgcc ttcggatctg 1500
133 tacaacgcgc tgccttacga tatgtaa 1527

```

136 <210> SEQ ID NO: 3

137 <211> LENGTH: 1549

138 <212> TYPE: DNA

139 <213> ORGANISM: Arabidopsis thaliana

141 <400> SEQUENCE: 3

```

142 gcaaggatcc atgtcgtggt ttctaatagc ggtggcgaca atcgccgcgc tcgtatccta 60
144 caagctaate caacgggctaa gatacaagtt cccaccaggc ccaagcccca agccgatcgt 120
146 cggtaacctc tacgacataa aaccgggtccg gttcagatgt tactacgagt gggctcaatc 180
148 ttatggacca atcatatcgg tctggatcgg ttcaattcta aacgtggtcg tatctagcgc 240
150 cgagctagca aaagaagttc tgaaagaaca cgaccagaaa ctcgccgacc ggcaccggaa 300
152 cagatcgacg gaagcattta gccgcaacgg tcaggatctt atatgggccc attatgggccc 360
154 tcattacgtg aaggtgagaa aagtttgcac gcttgagctc ttcacaccga aacgactcga 420
156 gtctctcaga cctatccgtg aagatgaagt caccgccatg gttgaatccg tcttcagaga 480
158 ctgtaacctt cctgaaaaca gagcaaaagg tttacaactg aggaagtact taggagcggg 540
160 tgcgttcaac aacataacgc ggctagcctt tgggaagcgt tttatgaacg ctgaagggtg 600
162 tgtggacgag caagggcttg agttcaaggc catagtatcc aacgggtctga agctaggtgc 660
164 ttcactgtca atagctgaac acatcccgtg gctcagggtg atgtttccgg ctgatgagaa 720
166 ggcgtttgct gagcacgggg ctcgctcgtg ccgcctcact cgagctatca tggaggagca 780
168 tactttggcc cgtcaaaaagt ctagtggagc gaaacagcat ttcgttgatg cgttgctaac 840
170 gttgaaggat cagtatgata ttagtgagga tactatcatt ggtcttctat gggatatgat 900
172 caccgcaggg atggacacga cagcgataac agcggaatgg gcgatggcgg aaatgatcaa 960
174 gaatccaaga gtgcaacaaa aagtgcaga agagttcgac agagtgggtg gacttgaccg 1020
176 gatcttaacc gaggcagatt tctcccgtt accttacttg caatgcgtgg tgaaagagtc 1080

```

RAW SEQUENCE LISTING

DATE: 08/23/2001

PATENT APPLICATION: US/09/931,267

TIME: 17:30:55

Input Set : A:\N1422005.txt

Output Set: N:\CRF3\08162001\I931267.raw

```

178 attcaggctg catcctccaa cgcctctaata gctacctcac cgaagcaacg cagatgtcaa 1140
180 gatcggaggc tatgatattc ccaaaggatc aaacgttcat gtgaatgtgt gggctgtggc 1200
182 tagagaccgc gctgtatgga aaaatccatt tgagtttaga ccagagagat tcttggaaga 1260
184 agatgttgac atgaagggtc atgatttttag gctgcttccg tttggagctg gaagacgggt 1320
186 ttgtcccgtg gcacaacttg gtatcaattt ggtaacttcg atgatgagtc atttgcttca 1380
188 ccattttgtt tggacacctc ctcaagggtc taaaccggag gagattgaca tgtctgaaaa 1440
190 ccctggactc gttacttaca tgcgtacccc tgtgcaagcg gttgcaacgc ctcggttgcc 1500
192 ttcggatctg tacaaacgcg tgccttacga tatgtaaatg aattcctga 1549
195 <210> SEQ ID NO: 4
196 <211> LENGTH: 508
197 <212> TYPE: PRT
198 <213> ORGANISM: Arabidopsis thaliana
200 <400> SEQUENCE: 4
202 Met Ser Trp Phe Leu Ile Ala Val Ala Thr Ile Ala Ala Val Val Ser
203 1 5 10 15
205 Tyr Lys Leu Ile Gln Arg Leu Arg Tyr Lys Phe Pro Pro Gly Pro Ser
206 20 25 30
208 Pro Lys Pro Ile Val Gly Asn Leu Tyr Asp Ile Lys Pro Val Arg Phe
209 35 40 45
211 Arg Cys Tyr Tyr Glu Trp Ala Gln Ser Tyr Gly Pro Ile Ile Ser Val
212 50 55 60
214 Trp Ile Gly Ser Ile Leu Asn Val Val Val Ser Ser Ala Glu Leu Ala
215 65 70 75 80
217 Lys Glu Val Leu Lys Glu His Asp Gln Lys Leu Ala Asp Arg His Arg
218 85 90 95
220 Asn Arg Ser Thr Glu Ala Phe Ser Arg Asn Gly Gln Asp Leu Ile Trp
221 100 105 110
223 Ala Asp Tyr Gly Pro His Tyr Val Lys Val Arg Lys Val Cys Thr Leu
224 115 120 125
226 Glu Leu Phe Thr Pro Lys Arg Leu Glu Ser Leu Arg Pro Ile Arg Glu
227 130 135 140
229 Asp Glu Val Thr Ala Met Val Glu Ser Val Phe Arg Asp Cys Asn Leu
230 145 150 155 160
232 Pro Glu Asn Arg Ala Lys Gly Leu Gln Leu Arg Lys Tyr Leu Gly Ala
233 165 170 175
235 Val Ala Phe Asn Asn Ile Thr Arg Leu Ala Phe Gly Lys Arg Phe Met
236 180 185 190
238 Asn Ala Glu Gly Val Val Asp Glu Gln Gly Leu Glu Phe Lys Ala Ile
239 195 200 205
241 Val Ser Asn Gly Leu Lys Leu Gly Ala Ser Leu Ser Ile Ala Glu His
242 210 215 220
244 Ile Pro Trp Leu Arg Trp Met Phe Pro Ala Asp Glu Lys Ala Phe Ala
245 225 230 235 240
247 Glu His Gly Ala Arg Arg Asp Arg Leu Thr Arg Ala Ile Met Glu Glu
248 245 250 255
250 His Thr Leu Ala Arg Gln Lys Ser Ser Gly Ala Lys Gln His Phe Val
251 260 265 270
253 Asp Ala Leu Leu Thr Leu Lys Asp Gln Tyr Asp Leu Ser Glu Asp Thr
254 275 280 285

```

RAW SEQUENCE LISTING

DATE: 08/23/2001

PATENT APPLICATION: US/09/931,267

TIME: 17:30:55

Input Set : A:\N1422005.txt

Output Set: N:\CRF3\08162001\I931267.raw

```

256 Ile Ile Gly Leu Leu Trp Asp Met Ile Thr Ala Gly Met Asp Thr Thr
257      290      295      300
259 Ala Ile Thr Ala Glu Trp Ala Met Ala Glu Met Ile Lys Asn Pro Arg
260 305      310      315      320
262 Val Gln Gln Lys Val Gln Glu Glu Phe Asp Arg Val Val Gly Leu Asp
263      325      330      335
265 Arg Ile Leu Thr Glu Ala Asp Phe Ser Arg Leu Pro Tyr Leu Gln Cys
266      340      345      350
268 Val Val Lys Glu Ser Phe Arg Leu His Pro Pro Thr Pro Leu Met Leu
269      355      360      365
271 Pro His Arg Ser Asn Ala Asp Val Lys Ile Gly Gly Tyr Asp Ile Pro
272      370      375      380
274 Lys Gly Ser Asn Val His Val Asn Val Trp Ala Val Ala Arg Asp Pro
275 385      390      395      400
277 Ala Val Trp Lys Asn Pro Phe Glu Phe Arg Pro Glu Arg Phe Leu Glu
278      405      410      415
280 Glu Asp Val Asp Met Lys Gly His Asp Phe Arg Leu Leu Pro Phe Gly
281      420      425      430
283 Ala Gly Arg Arg Val Cys Pro Gly Ala Gln Leu Gly Ile Asn Leu Val
284      435      440      445
286 Thr Ser Met Met Ser His Leu Leu His His Phe Val Trp Thr Pro Pro
287      450      455      460
289 Gln Gly Thr Lys Pro Glu Glu Ile Asp Met Ser Glu Asn Pro Gly Leu
290 465      470      475      480
292 Val Thr Tyr Met Arg Thr Pro Val Gln Ala Val Ala Thr Pro Arg Leu
293      485      490      495
295 Pro Ser Asp Leu Tyr Lys Arg Val Pro Tyr Asp Met
296      500      505
298 <210> SEQ ID NO: 5
299 <211> LENGTH: 508
300 <212> TYPE: PRT
301 <213> ORGANISM: Arabidopsis thaliana
303 <400> SEQUENCE: 5
305 Met Ser Trp Phe Leu Ile Ala Val Ala Thr Ile Ala Ala Val Val Ser
306 1      5      10      15
308 Tyr Lys Leu Ile Gln Arg Leu Arg Tyr Lys Phe Pro Pro Gly Pro Ser
309      20      25      30
311 Pro Lys Pro Ile Val Gly Asn Leu Tyr Asp Ile Lys Pro Val Arg Phe
312      35      40      45
314 Arg Cys Tyr Tyr Glu Trp Ala Gln Ser Tyr Gly Pro Ile Ile Ser Val
315      50      55      60
317 Trp Ile Gly Ser Ile Leu Asn Val Val Val Ser Ser Ala Glu Leu Ala
318 65      70      75      80
320 Lys Glu Val Leu Lys Glu His Asp Gln Lys Leu Ala Asp Arg His Arg
321      85      90      95
323 Asn Arg Ser Thr Glu Ala Phe Ser Arg Asn Gly Gln Asp Leu Ile Trp
324      100      105      110
326 Ala Asp Tyr Gly Pro His Tyr Val Lys Val Arg Lys Val Cys Thr Leu
327      115      120      125

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/931,267

DATE: 08/23/2001

TIME: 17:30:55

Input Set : A:\N1422005.txt

Output Set: N:\CRF3\08162001\I931267.raw

```

329 Glu Leu Phe Thr Pro Lys Arg Leu Glu Ser Leu Arg Pro Ile Arg Glu
330      130                      135                      140
332 Asp Glu Val Thr Ala Met Val Glu Ser Val Phe Arg Asp Cys Asn Leu
333 145                      150                      155                      160
335 Pro Glu Asn Arg Ala Lys Gly Leu Gln Leu Arg Lys Tyr Leu Gly Ala
336                      165                      170                      175
338 Val Ala Phe Asn Asn Ile Thr Arg Leu Ala Phe Gly Lys Arg Phe Met
339                      180                      185                      190
341 Asn Ala Glu Gly Val Val Asp Glu Gln Gly Leu Glu Phe Lys Ala Ile
342                      195                      200                      205
344 Val Ser Asn Gly Leu Lys Leu Gly Ala Ser Leu Ser Ile Ala Glu His
345      210                      215                      220
347 Ile Pro Trp Leu Arg Trp Met Phe Pro Ala Asp Glu Lys Ala Phe Ala
348 225                      230                      235                      240
350 Glu His Gly Ala Arg Arg Asp Arg Leu Thr Arg Ala Ile Met Glu Glu
351                      245                      250                      255
353 His Thr Leu Ala Arg Gln Lys Ser Ser Gly Ala Lys Gln His Phe Val
354                      260                      265                      270
356 Asp Ala Leu Leu Thr Leu Lys Asp Gln Tyr Asp Leu Ser Glu Asp Thr
357                      275                      280                      285
359 Ile Ile Gly Leu Leu Trp Asp Met Ile Thr Ala Gly Met Asp Thr Thr
360      290                      295                      300
362 Ala Ile Thr Ala Glu Trp Ala Met Ala Glu Met Ile Lys Asn Pro Arg
363 305                      310                      315                      320
365 Val Gln Gln Lys Val Gln Glu Glu Phe Asp Arg Val Val Gly Leu Asp
366                      325                      330                      335
368 Arg Ile Leu Thr Glu Ala Asp Phe Ser Arg Leu Pro Tyr Leu Gln Cys
369                      340                      345                      350
371 Val Val Lys Glu Ser Phe Arg Leu His Pro Pro Thr Pro Leu Met Leu
372      355                      360                      365
374 Pro His Arg Ser Asn Ala Asp Val Lys Ile Gly Gly Tyr Asp Ile Pro
375      370                      375                      380
377 Lys Gly Ser Asn Val His Val Asn Val Trp Ala Val Ala Arg Asp Pro
378 385                      390                      395                      400
380 Ala Val Trp Lys Asn Pro Phe Glu Phe Arg Pro Glu Arg Phe Leu Glu
381                      405                      410                      415
383 Glu Asp Val Asp Met Lys Gly His Asp Phe Arg Leu Leu Pro Phe Gly
384                      420                      425                      430
386 Ala Gly Arg Arg Val Cys Pro Gly Ala Gln Leu Asp Ile Asn Leu Val
387      435                      440                      445
389 Thr Ser Met Met Ser His Leu Leu His His Phe Val Trp Thr Pro Pro
390      450                      455                      460
392 Gln Gly Thr Lys Pro Glu Glu Ile Asp Met Ser Glu Asn Pro Gly Leu
393 465                      470                      475                      480
395 Val Thr Tyr Met Arg Thr Pro Val Gln Ala Val Ala Thr Pro Arg Leu
396                      485                      490                      495
398 Pro Ser Asp Leu Tyr Lys Arg Val Pro Tyr Asp Met
399                      500                      505
401 <210> SEQ ID NO: 6

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/931,267

DATE: 08/23/2001

TIME: 17:30:56

Input Set : A:\N1422005.txt

Output Set: N:\CRF3\08162001\I931267.raw

L:11 M:270 C: Current Application Number differs, Replaced Current Application No

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date